The Masked Politic

Joseph Marx and Steven Yan

Business Proposal

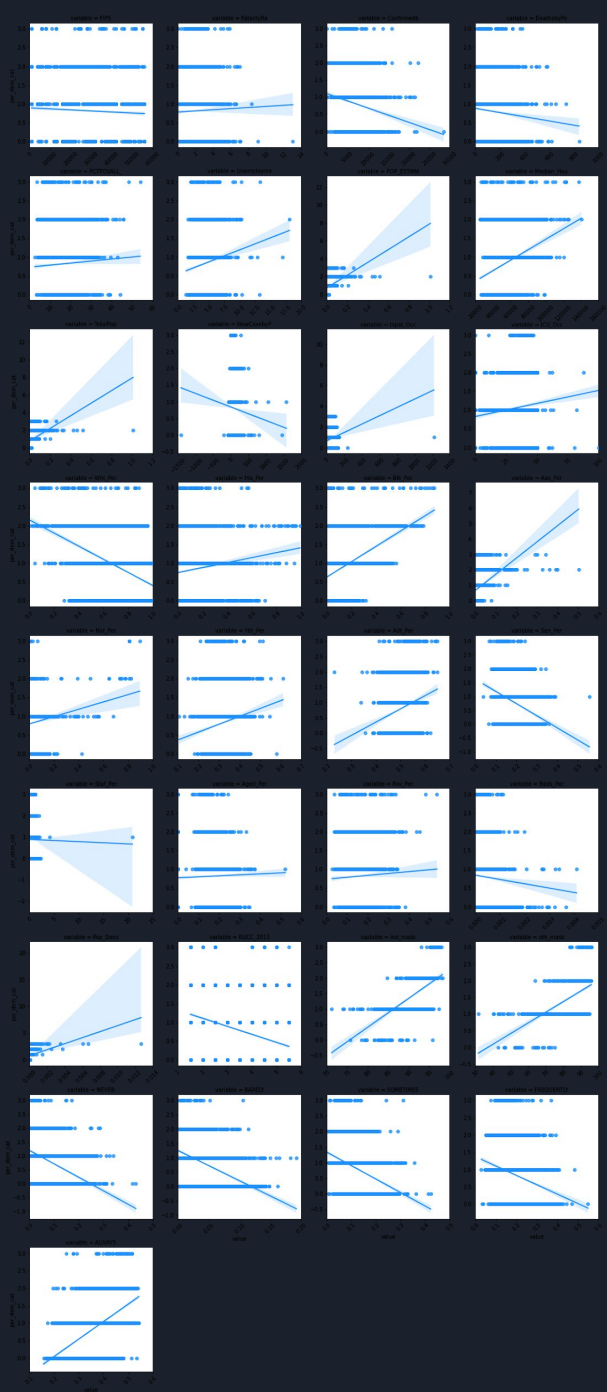
- Political polling has fallen short
- It’s becoming increasingly difficult to reach representative voter samples
- Fund a larger scale survey pipeline for COVID mask compliance

Data Collection

- Survey conducted by Carnegie Mellon on Mask Wearing
- Survey conducted by New York Times on Mask Wearing
- Database of COVID maintained by Johns Hopkins
- Election Data from Kaggle

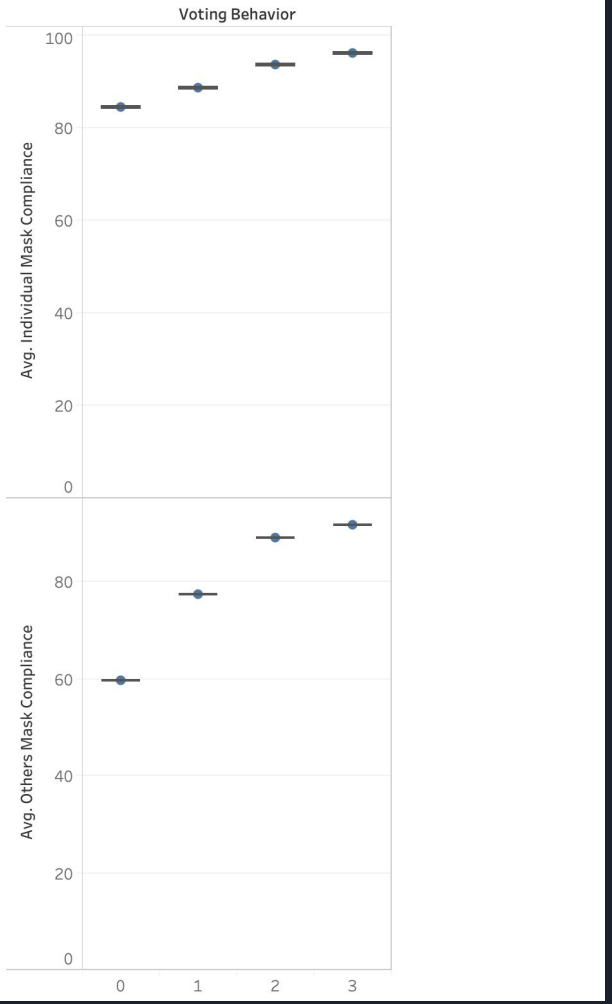
EDA

Explore relationship between features and voting behavior



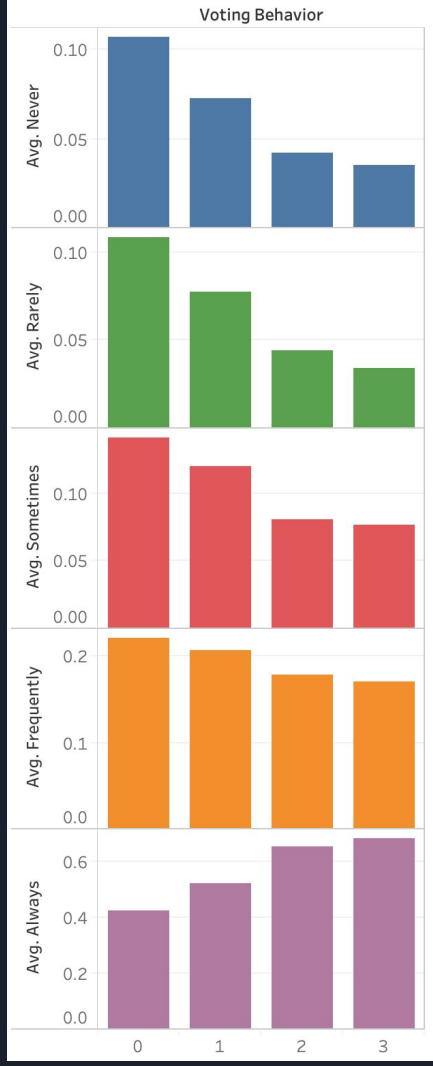
Carnegie Mellon Data

Mask Compliance Percentage vs. Voting Behavior



New York Times Study

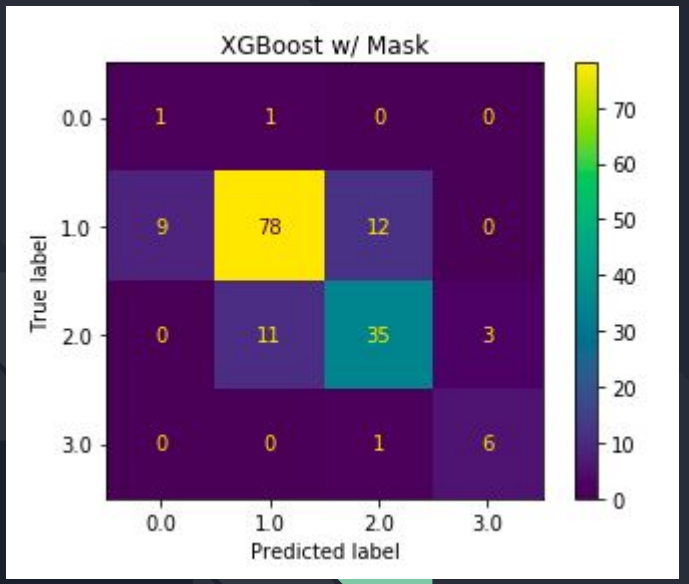
Mask Wearing Behavior vs. Voting Behavior



Modeling

- Train/Test Split, Scaling, SMOTE
- Created functions for quick model evaluation
- Parallel tests on datasets with and without features on mask compliance
- Different classifier models used with GridsearchCV:
  - Random Forest
  - XGBoost
  - Logistic Regression
- Each model was evaluated with a weighted-F1, Accuracy, and ROC/AUC Score

Results



	F1 Score		Accuracy		ROC/AUC	
Random Forest	0.552	0.692	0.497	0.662	0.733	0.838
XGBoost	0.711	0.781	0.682	0.764	0.801	0.889
Logistic Regression	0.599	0.716	0.541	0.682	0.739	0.82

Analysis

Models are weakest on predicting far left and far right counties

There is a boost in model performance for all tested metrics on all tested models by adding mask feautres

XGBoost was the strongest performing of our models

It is our recommendation that mask data be used in a model predicting on the 2020 U.S. Presidential Election

Next Steps

Test on future presidential/midterm elections

Utilize effective imputation methods

Study changing politicization of the pandemic

Model with NYT data to show the modeling value of using all counties

Secure job with Nate Silver

Thank You!

jmarx@hash.fyi

stevenyan@uchicago.edu